BAB I

PENDAHULUAN

1.1 Latar Belakang

Saat ini jumlah pengguna Linux di Indonesia tidak sedikit lagi seperti pada tahun 90-an. Tercatat pada tanggal 3 April 2010, pengguna distro Linux Ubuntu yang terdaftar pada komunitas Ubuntu Indonesia terdapat 5027 orang, belum yang tidak terdaftar dan juga pengguna distro Linux yang lainnya. Untuk pengguna Linux yang terdaftar di daerah Bandung saja pada tanggal 3 April 2010 mencapai 526 orang dan di daerah Bogor terdapat 556 orang. Banyaknya pengguna Linux saat ini membuat seseirang yang ingin belajar Linux menjadi lebih mudah.

Salah satu media para pengguna Linux berkumpul untuk berbagi informasi ataupun mencari solusi adalah grup milis. Banyak sekali grup milis di Indonesia seperti Ubuntu Indonesia, Klub Linux Bandung, KPLI (Komunitas Pengguna Linux Indonesia) Bogor dan yang lainnya. Salah satu grup milis yang sangat aktif adalah Ubuntu Indonesia. Setiap bulannya tulisan yang masuk ke milis Ubuntu Indonesia rata-rata mencapai 1000 sampai 2000 pesan, bahkan pada bulan Juli 2007 sempat mencapai 3436 pesan yang masuk ke milis. Banyaknya informasi yang ada di milis ini dapat dijadikan sumber orang yang ingin belajar Linux.

Permasalahan muncul ketika seseorang ingin mencari informasi atau mencari solusi dari permasalahan yang didapat. Orang tersebut harus mencari dari satu milis ke milis yang lain untuk mendapatkan solusi yang tepat dengan permasalahan yang dimilikinya. Hal ini membuat proses pencarian solusi atau

informasi menjadi lama dan tidak efektif, karena orang tersebut harus mencarinya dari satu milis ke milis yang lain, sedangkan banyak sekali milis yang ada di Indonesia, sehingga perlu adanya media yang dapat mencarikan solusi permasalahan orang tersebut dari banyaknya milis yang ada, salah satunya adalah *Search Engine*.

Sampai saat dibuatnya penelitian ini, belum ada media yang dapat digunakan sebagai sistem pencari informasi untuk seluruh grup milis di Indonesia, khususnya milis Linux. Oleh sebab itu perdi dibuat sebuah sistem pencarian untuk grup milis, dalam kasus ini adalah sistm pencarian grup milis Linux di Indonesia. Namun yang perlu diperhatikan untuk membangun sistem pencarian untuk grup milis Linux di Indonesia adalah, seberapa besarkah data elektronik yang terdapat pada milis Linux di Indonesia? Mengapa perlu tahu? Hal ini dikarenakan sistem pencarian yang dibangun haruslah dapat melakukan proses pengolahan (baik itu proses *indexing* atau proses pencarian) pada data dengan ukuran besar. Tercatat pada tanggal 3 April 2010 jam 12:20, terdapat 74.345 tulisan yang ada di milis Ubuntu Indonesia, 17.828 tulisan di milis KPLI Bogor dan 11048 tulisan di milis Klub Linux Bandung. Jika diasumsikan terdapat 1 GB data pada setiap milis Linux yang ada di Indonesia, maka untuk 10 milis saja sudah mencapai 10 GB. Hal ini berarti sistem pencarian yang akan dibangun harus bisa melakukan proses pengolahan pada data yang berukuran besar dengan cepat.

Masalah dari data elektronik yang besar tersebut sangaat sederhana: sementara kapasistas media penyimpanan (*hard drive*) selalu meningkat dengan cepat dari waktu ke waktu, namun kecepatan rata-rata data yang dapat dibaca oleh

media penyimpanan tersebut tidak mengimbangi peningkatan kapasitasnya. Tom White, dalam bukunya *Hadoop: The Definitive Guide* mengatakan, "Sebuah *hard drive* tahun 1990 memiliki kapasistas 1370 MB dan kecepatan transfer sekitar 4,4 MB/s, sehingga dapat membaca keseluruhan data hanya dalam kurun waktu lima menit. Hampir 20 tahun setelah itu, mundul hard drive dengan kapasistas 1 TB, namun kecepatan transfer sekitar 100 MB/s, sehingga membutuhkan waktu dua setengah jam untuk membaca keseluruhan data." Dua setengah jam adalah waktu yang sangat lama untuk melakukan proses membaca, apalagi jika melakukan proses menulis data. Untuk menangani masalah tersebut Tom White memberi solusi, "Hal yang dapat kita lakukan untuk mengurangi waktu membaca data dari *hard drive* adalah dengan membacanya dari banyak *hard drive* secara bersamaan. Bayangkan jika kita memiliki 100 *hard drive* dan setiap *hard drive* menyimpan 100 MB, jika dikerjakan secara paralel, maka kita dapat membaca keseluruhan date tersebut dalam waktu kurang dari dua menit."

MapReduce dikemukakan oleh Jeffrey Dean dan Sanjay Ghemawat tahun 2004 melalui publikasinya yang berjudul *MapReduce: Simplied Data Processing on Large Clusters* melalui Google Labs. Dalam publikasinya, mereka mengatakan bahwa *MapReduce* merupakan model pemrograman yang diimplementasikan untuk melakukan pemrosesan atau pengolahan data elektronik dalam jumlah besar. Aplikasi yang dibuat menggunakan model pemrograman *MapReduce* akan secara otomatis paralel, hal ini membuat pengembang yang tidak berpengalaman terhadap komputasi paralel dam sistem terdistribusi akan mudah membangun aplikasi pada sistem terdistribusi yang besar.

Dari permasalahan diatas ternyata sangat sulit untuk membangun sistem pencarian untuk milis Linux di Indonesia dikarenakan besarnya data elektronik yang ada. Oleh sebab itu perlu diterapkan model pemrograman *MapReduce* pada sistem pencarian yang akan dibangun.

1.2 Perumusan Masalah

Dari latar belakang ditemukan beberapa masalah yang dirumuskan dalam satu rumusan masalah yaitu bagaimana menerapkan model pemrograman *MapReduce* pada sistem pencarian grup milis Linux di Indonesia.

1.3 Maksud dan Tujuan

1.3.1 Maksud

Pembuatan penelitian ini membangun sistem pencarian untuk milis Linux di Indonesia dengan menerapkan model pemrograman *MapReduce*.

1.3.2 danTujuan

Tujuan dari pembangunan sistem pencarian untuk milis Linux di Indonesia dengan menerapkan model pemrograman *MapReduce* adalah untuk mempermudah pengguna Linux di Indonesia dalam mencari informasi yang ada dalam milis-milis Linux secara cepat dan mudah.

1.4 Batasan Masalah

Adapun ruang lingkup yang akan dibahas sangatlah luas, untuk itu diperlukan batasan masalah sebagai berikut :

1. Sistem yang akan dibangun berjalan pada sebuah komputer bukan *cluster*.
2. Sistem yang akan dibangun adalah sistem pencarian untuk grup milis Linux di Indonesia

3. Sistem hanya mengelola milis yang terdapat di Google Groups dan Yahoo Groups.
4. Model analisis sistem adalah pemodelan berorientasi objek dan peralatan yang digunakan adalah UML (Unified Modeling Language) versi 2.
5. Sistem akan dibangun berbasiskan *website* dan berjalan pada *webserver* Glassfish versi 3.
6. Sistem akan dibangun menggunakan bahasa pemrograman Java.
7. Sistem akan berjalan pada sistem operasi Ubuntu dan Windows.

1.5 Metodologi Penelitian

1.5.1 Pengumpulan Data

Metodologi Penelitian yang digunakan dalam pengumpulan data yang berkaitan dengan penyusunan skripsi dan pembuatan sistem pencarian ini adalah sebagai berikut.

1. Studi kepustakaan.

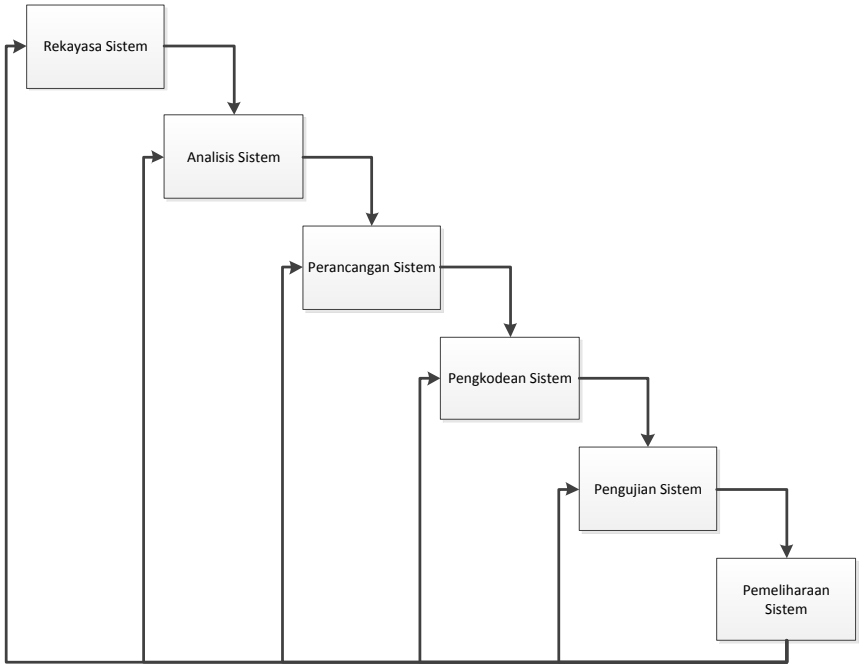
   Tahap ini digunakan untuk mencari informasi yang berhubungan dengan permasalahan yang akan dibahas dengan bersuber pada buku-buku serta bacaan lain yang kiranya dapat mebantu menyelesaikan pembangunan aplikasi.

2. Studi dokumentasi.

   Dengan mengumpulkan informasi mengenai data yang diperlukan agar lebih mudah untuk didefinisikan dan dirumuskan pada permasalahan yang ada.

1.5.2 Pembangunan Perangkat Lunak

Pembangunan sistem pencarian ini menggunakan metodologi Waterfall dengan gambar 1.1 sebagai berikut



Gambar 1.1 *Waterfall*

1. Pekayasa Sistem

   Tahapan ini merupakan kegianan pengumpulan data sebagai pendukung pembangunan sistem serta menentukan ke arah mana aplikasi ini akan dibangun.

2. Analisis Sistem

   Mengumpulkan kebutuhan secara lengkap kemudian dianalisis dan didefinisikan kebutuhan yang harus dipenuhi oleh aplikasi yang akan dibangun.

3. Perancangan Sistem

Perancangan antarmuka dari hasil analisis kebutuhan yang telah selesai dikumpulkan secara lengkap.

4. Pengkodenan Sistem

   Hasil perancangan sisten diterjemahkan ke dalam kode-kode dengan menggunakan bahasa pemrograman yang telah ditentukan,

5. Pengujian Sistem

   Penyatuan unit-unit program kemudian diuji secara keseluruhan.

6. Pemeliharaan Sistem

   Mengoperasikan sistem dilingkungannya dan melakukan pemeliharaan, seperti penyesuaian atau perubahan karena adaptasi dengan situasi yang sebenarnya.

1.5.3 Sistematika Penulisan

Sistematia penulisan laporan ini dibagi dalam beberapa bab dengan pokok pembahasan secara umum sebagai berikut :

BAB I PENDAHULUAN

Bab ini berisikan latar belakang, perumusan masalah, maksud dan tujuan, batasan masalah, metodologi penelitian dan sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini berisikan landasan-landasan teori yang digunakan dalam pembangunan sistem.

BAB III ANALISIS DAN PERANCANGAN SISTEM

Bab ini berisi analisis kebutuhan dalam membangun aplikasi ini, analisis sistem yang sedang berjalan pada aplikasi ini sesuai dengan metode pembangunan

perangkat lunak yang digunakan. Selain itu terdapat juga perancangan antarmuka untuk aplikasi yang akan dibangun sesuai dengan hasil analisis yang telah dibuat.

BAB IV IMPLEMENTASI DAN PENGUJIAN

Bab ini berisi hasil implementasi dari hasil analisis dan perancangan yang telah dibuat disertai juga dengan hasil pengujian dari aplikasi ini.

BAB V KESIMPULAN DAN SARAN

Bab ini berisi kesimpun tentang keseluruhan dari pembangunan aplikasi ini dan saran tentang aplikasi ini untuk masa yang akan datang.